DOCUMENT RESUME

ED 441 015 TM 030 813

AUTHOR Fuller, Michael L.
TITLE Teacher Judgment as Formative and Predictive Assessment of Student Performance on Ohio's Fourth and Sixth Grade Proficiency Tests.
PUB DATE 2000-04-00
NOTE 45p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE Numerical/Quantitative Data (110) -- Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Education; *Elementary School Teachers; Formative Evaluation; *Prediction; Standardized Tests; *Student Evaluation; *Teacher Attitudes; *Test Results
IDENTIFIERS Ohio Fourth Grade Proficiency Test; *Ohio Sixth Grade Proficiency Test

ABSTRACT

Ninety teachers in grades three through six were asked to judge the likelihood of their students' passing Ohio's Fourth or Sixth Grade Proficiency Tests. Judgment ratings consisted of "likely to pass," "uncertain to pass," or "unlikely to pass." These ratings were collected in January 1998, 3 months prior to the administration of the proficiency tests. Test results were collected the following June. In general, third- and fourth-grade teachers were more accurate in identifying those students who passed than those who failed. Fifth- and sixth-grade teachers were mixed in their judgments. Regardless of teacher grade level, students judged likely to pass had higher mean proficiency scores than those judged uncertain or unlikely to pass. No significant differences were found in teachers' judgments in high-performing schools and low-performing schools. These results show that teacher judgment can serve as a predictive assessment for likely performance on Ohio's Fourth or Sixth Grade Proficiency Tests. Preliminary results are also presented for using teacher judgment as a formative assessment. (Contains 22 tables and 30 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 441 015

Running Head: TEACHER JUDGMENT

Teacher Judgment as Formative and Predictive Assessment of Student Performance on Ohio's Fourth and Sixth Grade Proficiency Tests

Michael L. Fuller, PhD
Muskingum Valley Educational Service Center
205 North 7th Street
Zanesville, Ohio 43701
(740) 452-4518
mfuller@mvesc.k12.oh.us

M. Fuller

In-Works Paper

March 30, 2000

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

A paper presented at the American Educational Research Association Annual Meeting, New Orleans, April, 2000. A note of appreciation is extended to Eric Pickerington who, during his school psychology internship year, assisted in collecting data and conducting analyses for this study.

TM030813

BEST COPY AVAILABLE

Abstract

Ninety (90) teachers in grades three through six were asked to judge the likelihood of their students passing Ohio's Fourth or Sixth Grade Proficiency Tests. Judgment ratings consisted of likely to pass, uncertain to pass, or unlikely to pass, and were collected in January, 1998, three months prior to administration of the proficiency tests. Test results were collected the following June. Generally, third and fourth grade teachers were more accurate in identifying those students who passed than those who failed. Fifth and sixth grade teachers were mixed in their judgments. Regardless of teacher grade level, students judged likely to pass had higher mean proficiency scores than those judged uncertain or unlikely to pass. No significant differences were found in teachers' judgments in high performing schools and in low performing schools. These results show that teacher judgment can serve as a predictive assessment for likely performance on Ohio's Fourth or Sixth Grade Proficiency Tests. Preliminary results are also presented for using teacher judgment as formative assessment.

Teacher Judgment as Formative and Predictive Assessment of Student Performance on Ohio's Fourth and Sixth Grade Proficiency Tests

Although evidence can be cited to the contrary (e.g., Robinson & Brandon, 1992; Carson, Huelskamp & Woodall, 1993; Berliner & Biddle, 1995; Bracey, 1997; McQuillan, 1998; Levin, 1998; Forgione, 1998), it is widely accepted that our public schools are in need of reform. For example, in publications such as Quality Counts: A Report Card on The Condition of Public Education in the 50 States (1997), we find such quotes as "Despite 15 years of earnest efforts to improve public schools and raise student achievement, states haven't made much progress." (p.3) In Quality Counts'99, the editorial opens with, "The pressure is on. After years of exhorting and cajoling schools to improve, policymakers have decided to get tough." (p.5)

Accountability is now the central feature of educational reform. Accountability grew out of the standards-based reform movement of the 1990's. Standards prescribe what students should know and be able to do. Assessments, linked to the standards, are used to determine whether schools and students are meeting the standards. Forty-nine states have or are developing common academic standards for their students (American Federation of Teachers, 1997). Forty-eight states test their students and 36 publish annual report cards on individual schools (Quality Counts'99, 1999). Like never before schools are accountable for results. And in the main, student performances on high stakes tests are the results.

High stakes refer to the important consequences these tests hold for schools (e.g., public ratings based on student achievement) and for students (e.g., promotion and graduation). The National Association of State Boards of Education has gone on record that state assessments of student achievement have consequences for students who take them and for the schools that give them (Education Week, October 22, 1997). Indeed, such pervasive use of tests for accountability purposes prompted the US Congress to order a study of high stakes testing (Heubert & Hauser, 1999). The study's mandate was to determine whether tests are used in an appropriate and nondiscriminatory manner, and whether they adequately assess reading and mathematics in a manner likely to yield accurate information related to these achievement skills.

Ohio's Efforts

Ohio is serious about educational reform. Ohio has developed model competency-based programs in language arts, mathematics, social studies and science. Making Standards Matter

(American Federation of Teachers, 1997) rated Ohio's math standard as exemplary and its English standard as strong. In Quality Counts '98 Ohio was given an A- in the area of student and assessments standards (6th highest in the nation) and a B (4th highest in the nation) in the area of teachers who have the knowledge and skills to teach to higher standards. However, in Quality Counts 2000, Ohio is now ranked 21st in standards and assessment and 17th in teacher quality. Central to Ohio's improvement efforts are the Ohio's Proficiency Tests, a series of high stakes tests.

The Ohio Proficiency Tests were enacted into law in 1987 (Ohio Department of Education, 1996). The Ninth Grade Proficiency Tests were first administered to the freshman class of 1990. Proficiency tests now exist in grades 4, 6, 9, and 12. The Ninth Grade Proficiency Tests will be replaced by new high school graduation qualifying exams, starting in the 2002-2003 school year. Passing the Ninth Grade Tests is now necessary for a high school diploma. After phase out of the Ninth Grade Tests, passing the new high school graduation qualifying exams will be a requirement for high school graduation.

In May, 1999 99% of Ohio's twelfth graders had passed all parts of the Ninth Grade Proficiency Tests and were therefore eligible for high school graduation. The 1% who have yet to pass all parts corresponds to 2,561 twelfth graders who did not graduate on time. Obviously, the Ninth Grade Proficiency Tests and the new high school qualifying exams are high stakes. However, the other proficiency tests have high stake consequences as well.

Amended Substitute Senate Bill 55 (SB 55) was passed by Ohio's General Assembly in the fall of 1997 and was intended to serve as the carrier for all educational reform efforts in Ohio. SB 55 requires fourth graders, starting in the Fall of 2001, to pass the reading portion of the Fourth Grade Ohio Proficiency Tests in order to be promoted to the fifth grade. SB 55 calls for 3 opportunities to pass the reading test, but it is clear some students will not be promoted. Although it is impossible to say with certainty how many children will be affected, we can get some rough idea from the most recent Fourth Grade Proficiency Tests administration. In March, 1999 60% of Ohio's fourth graders passed the reading test. Were this 2001, 40%, or 51,199 students, would have to retake the reading test. If 75% of the students who failed the reading test eventually passed, Ohio would still have to retain nearly 13,000 fourth graders.

Included in SB 55 was the requirement to rate public schools on the basis of 18 performance indicators, 16 (89%) of which relate to proficiency performances. In July, 1999

House Bill 282 increased the number of performance indicators from the 18 in SB 55 to 27. Proficiency test results now make up 25 (93%) of the performance indicators. The remaining two indicators relate to student attendance and high school graduation rates.

Starting in February 2000, Ohio's school districts will be placed in one of four designations: 1) effective, meets 26 or 27 indicators; 2) continuous improvement, meets 14-25 indictors; 3) academic watch, meets 9-13 indicators; and 4) academic emergency, meets 0-8 indictors. Each district's designation will be contained in a report card that will be disseminated to the public. These designations are intended to serve as broad benchmarks of the quality of education available in the schools.

Based on 1999 data, less than 5% of districts fall in the effective category, 62% are in continuous improvement, and over 21% and 11% respectively are in academic watch or emergency. Except for effective schools, all others must develop continuous improvement plans to move into the effective category. Those in continuous improvement have 5 years, in academic watch 8 years, and those in emergency have 13 years to do so. In addition, minimal yearly progress gains must be met. Those districts failing to make necessary gains are subject to various state interventions.

Assessment Issues

It is easy to see that Ohio's Proficiency Tests are high stakes. These tests are summative; passing performance is based on predetermined criteria (Ohio Department of Education, April, 1997). Each test is based on learning outcomes adopted by the State Board of Education. Learning outcomes define the proficiencies students are expected to possess and apply as a result of their accumulated educational experiences. Each proficiency test is composed of 5 subtests: writing, reading, mathematics, citizenship, and science. Each subtest consists of strands and related outcomes. For example, writing has 9 outcomes grouped into 4 strands. One of the writing strands is content. Content is measured by 2 outcomes: 1) a response that stays on topic; and 2) the use of details to support the topic. Outcomes vary in number and kind depending on the grade of the proficiency test. The Fourth Grade Reading Proficiency Test has 4 strands and 20 learning outcomes (Ohio Department of Education, 1995, August). The Sixth Grade Reading Proficiency Test has 4 strands and 18 learning outcomes (Ohio Department of Education, 1995, August).

For reading, mathematics, citizenship, and science, raw scores are converted to scaled scores. For writing, a rubric scoring system is used to assign scores. On the Fourth Grade Reading Proficiency Test, a student must earn a raw score of 34 out of 42 to pass (Ohio Department of Education, Assessment Center, June, 1999). This converts to a scaled score of 217. On the writing test, the student must earn a rubric score of 5 out of 8 to pass. Each year proficiency tests are independently evaluated for psychometric adequacy.

Given that proficiency tests are summative and given infrequently, various interim measures are needed to serve formative and predictive functions. It is an understatement to say that work is needed in this area.

Competency-based education. Competency-based education (CBE) has been required in Ohio's public schools since 1983 (Ohio Department of Education, 1995). Schools are required to develop criterion-based instructional and performance objectives for all academic disciplines and to develop assessment strategies and methods to judge whether satisfactory learner progress is occurring. Furthermore the instructional and performance objectives and related assessments are expected to support the outcomes of proficiency tests. However, school districts are free to develop their own CBE programs, independent of the model programs adopted by the State Board of Education. As a consequence, most assessments were designed to reflect minimal student performance, with little consideration given to the psychometric integrity of these locally developed assessments. In a study by Loe and Fuller (1997), 99% of students, in one rural elementary school, who passed the Fourth Grade Mathematics Proficiency Test passed their third grade mathematics CBE evaluations. However, 92% of those who failed the Fourth Grade Mathematics Proficiency Test passed their third grade CBE evaluations. Although limited to one school district, these findings have important implications and warrant further investigation. It appears that current CBE evaluations may be neither instructionally relevant nor predictive.

Off-year proficiency-based assessments. Most school districts in Ohio purchase commercially produced off-year assessments. These assessments have the "look and feel" of proficiency tests. The format is comparable to what students will experience on the proficiency tests and the questions asked appear to sample the proficiency outcomes. Like proficiency tests, off-year measures use a criterion to determine passing performance. However the manner in which the criterion is derived for passing and failing is questionable. Passing scores are based on the rank order of the previous year's proficiency scores. For example, if last year 40% of a

district's fourth graders passed the reading portion of the proficiency tests, 40% would be used for this year's cut scores. Let's say that third graders are administered an off-year proficiency test. Their performances on the off-year test would then be rank ordered. For reading, third grade students would pass if their performance fell in the upper 40% of their rank ordered distribution. However, this approach may mean that students can pass 80-to-90% of the items on a test and still fail, something schools find difficult to explain to parents. Apart from glossy promotional materials, no technical studies of the reliability and validity of these off-year tests are offered to school districts.

Norm-referenced assessments. Given the requirements for CBE and proficiency testing in Ohio schools, interest in and use of standardized norm-referenced testing has declined. Many question the relevance of such tests, particularly given the reliance on multiple choice questions, and the additional expense associated with purchasing and scoring these tests. However, unlike CBE evaluations and off-year proficiency assessments, norm-referenced assessments are usually technically sound instruments. Not only can these tests be used to compare current performance from a normative perspective, they can also be used to make predictions of future performances on dissimilar measures.

Fuller and DeMarie-Dreblow (1992) found that students' performance on standardized achievement tests in their fourth grade year highly correlated to their performance on the Ninth Grade Proficiency Tests some five years later. Logistic regression models were estimated that correctly identified 84% of students who failed and 76% who passed the Ninth Grade Mathematics Proficiency Test. The models also correctly identified 30% who failed and 96% who passed reading, and 18% who failed and 95% who passed writing.

Correlational analyses now exist between the Stanford Achievement Test-9 (SAT-9, 1999) and the Ohio Proficiency Tests. For example, SAT-9 Total Reading correlates .783 with the Ninth Grade Reading Proficiency Test. SAT-9 Total Math correlates .869 with Ninth Grade Mathematics, and SAT-9 Total Social Science correlates .783 with Ninth Grade Citizenship.

Teacher judgment. Although norm-referenced measures have distinct advantages over current CBE and off-year proficiency measures, all three share a number of limitations. They are costly in money and time, and are administered infrequently. Off-year proficiency and norm-referenced assessments have once a year administration cycles. CBE's are given no more than two-to-three times a year. Off-year proficiency and norm-referenced assessments involve a

lengthy turnaround time for the scoring and preparation of student reports. All three types of measures are summative; they represent a final test of students' knowledge and skills. Consequently they can not provide rapid and repeatable assessment of student progress, a formative assessment function. An area offering some promise as an adjunct to formative and predictive assessment is teacher judgment.

In a study by Demaray and Elliot (1998), teachers' judgments of students' academic achievement were reported to be accurate and could be gained through a rating scale format. They also demonstrated that teachers' direct judgment of students' item-by-item performances on the Kaufman Test of Educational Achievement was highly related to the students' actual performances. In a study by Hartman and Fuller (1997), teachers' rank ordering of their students' reading skills was shown to be highly correlated (.81 to .97) with students' subsequent performance on curriculum-based measures of reading and on the Word and Comprehension sections of the Stanford Achievement Test.

We see then that teachers can accurately judge student performances on achievement measures, at least under certain circumstances. To date, no reports on the ability of teachers to judge student performances on proficiency tests exist. Yet, this line of research is important for several reasons. If teachers can accurately make judgments of proficiency results, then this information can be used to convey an important likely future student status, that is, pass or fail. In turn, instructional and other support resources can be differentially allocated based on students' likelihood of passing or failing. Teacher judgment in this context serves a predictive function. Teacher judgment also can serve a formative function. Teacher judgment has the potential to be used as a rapid, repeatable, and inexpensive means of monitoring student progress, which in turn can assist teachers in grouping students for instructional purposes.

In this paper I present evidence of teachers' ability to predict the likelihood of students' passing or failing proficiency tests. I also include preliminary findings for the use of judgments as formative assessment.

Method

In January, 1998 teacher ratings from 23 schools (4 districts) were collected of students' likelihood of passing the Fourth or Sixth Grade Proficiency Tests. Ninety (90) teachers in grades three through six completed judgments on 2,476 students. A judgment rating sheet, which was developed by this author, was given to each teacher. A copy of the judgment sheet is in the

Appendix. Teachers were asked to print each of their student's names on the sheet and then to circle whether the student was likely to pass, uncertain to pass, or unlikely to pass each of the five subtests of the Fourth or Sixth Grade Proficiency Tests. Teachers were asked not to do any additional testing to arrive at a judgment, but to rate a student on the basis of the teacher's current knowledge of that student. Fourth grade teachers rated the likelihood of their students passing the Fourth Grade Proficiency Tests in March, 1998. Third grade teachers were asked to judge the likelihood of the students they had in the 1996-97 school year, who would now be fourth graders, to pass the Fourth Grade Proficiency Tests. That is, third grade teachers were asked to print the names of the third graders they had the previous year and to judge how likely each was to pass the Fourth Grade Proficiency Tests. In this fashion, many of the same students were rated by both third and fourth grade teachers. The same procedure was followed for fifth and sixth grade teachers. Sixth grade teachers judged their sixth grade students' likelihood of passing the Sixth Grade Proficiency Tests. Fifth grade teachers judged the likelihood of students they had the previous year to pass the Sixth Grade Proficiency Tests. In schools where team teaching occurred, the teachers pooled their judgment for each student. Each student's proficiency test results were collected in the Summer of 1998.

Results

Accuracy. The first series of analyses consisted of the accuracy of teacher judgments. Tables 1 through 10 list the judgments of third and fourth grade teachers in one school and fifth and sixth grade teachers in another school. Both schools were from the same school district. Given the number of tables associated with listing teacher accuracy, for this paper, I decided to limit presentation of this part of the analyses to these two schools. Results related to the accuracy for all schools will be presented in a summary fashion in Tables 11 and 12.

In Table 1, the accuracy of third and fourth grade teachers' judgments of their students' likelihood of passing the Fourth Grade Writing Proficiency Test is presented. A pass in writing is a rubric score of 5 or 6. Advanced pass is 7 or 8, and fail is 0 to 4. Writing judgments and writing proficiency results were collected for 96 fourth graders in this school. Eighty-five of those 96 students were also rated by third grade teachers in that school.

Of the 61 students who passed writing, slightly more than 80% were judged by their fourth grade teachers as likely to pass. Fifty-four of those 61 students were also judged by their third grade teachers. In this case, nearly 54% were identified as likely to pass. Fourth grade

teachers were uncertain about 16% of those who passed writing, and judged 3% of those who passed as unlikely to pass. Third grade teachers were uncertain about 33% who passed, and judged 13% of those who passed as unlikely to pass. Fourth grade teachers judged as likely to pass all students who earned an advanced pass. Third grade teachers judged 60% of those passed at the advanced level as likely to pass. No students who passed at the advanced level were judged as unlikely to pass.

Teachers were much less accurate in their judgments of those who failed the writing test. Of the 27 who failed, fourth grade teachers judged approximately 19% as unlikely to pass. Slightly more than 44% of those who failed were judged as likely to pass and the remaining 37% were judged uncertain to pass. Third grade teachers were more accurate, in that they judged 35% of the students they had as unlikely to pass.

Insert Table 1 About Here

For these third and fourth grade teachers, their best judgment was related to student performance on the Fourth Grade Citizenship Test. Table 2 presents these results. A scaled score of 208 to 249 equated to a pass for citizenship. A score of 250 or greater represented an advanced pass. Fourth grade teachers judged as likely to pass 86% of those who actually passed. Of those who passed, fourth grade teachers only judged 1% as unlikely to pass. Sixty-nine (69) of the 80 students who passed were also judged by their third grade teachers. In this case, third grade teachers judged 72% of the 69 students as likely to pass. Only 4% of the students who passed were judged as unlikely to pass. Third and fourth grade teachers, respectively, correctly judged 89% and 100% of the students who passed at the advanced level. For those who failed, fourth grade teachers judged a little more than 33% as unlikely to pass. Third grade teachers judged nearly 43% of those who failed as unlikely to pass.

Insert Table 2 About Here

Third and fourth grade teachers consistently were more accurate in their judgments of those who passed or advance passed than those who failed. Third and fourth grade teacher judgments for reading, math, and science are shown in Tables 3 to 5.

Insert Tables 3-to-5 About Here

Fifth and sixth grade teachers were often similarly accurate in judging those likely to pass and those unlikely to pass, with the exceptions of sixth grade teachers' judgments of students' performances in writing and science. In these instances, sixth grade teachers judged as likely to pass 19% of the 59 students who passed writing, and 38% of the 96 students who passed science. However 71% and 63% of those failing writing and science, respectively, were correctly judged as unlikely to pass. Only about one third of students judged by sixth grade teachers were judged by fifth grade teachers. Not all of the fifth grade teachers completed judgment ratings. These findings are shown in Tables 6-to-10.

Insert Tables 6-to-10 About Here

Table 11 shows the difference in accuracy in correctly judging fourth grade student performance in all 23 schools. Median percent correct judgment of students passing writing, reading, and mathematics was significantly greater, based on the Wilcoxin Sign Test, than percent correct judgment of students failing those respective areas. No significant differences were found between correct judgment of passing and failing for citizenship and science.

Insert Table 11 About Here

Schools giving the Fourth Grade Proficiency Tests were then rank ordered according to students' performances. Judgments of fourth grade teachers in schools falling in the first quartile were compared to the judgments of teachers in schools in the fourth quartile. Relative to the performance of the 23 schools in this study on proficiency tests, schools within the first quartile of performance can be characterized as low performers; those in the fourth quartile as high

performers. Based on the Mann-Whitney U Test, no significant differences were found between high and low performing schools in teachers' correctly judging students passing. As seen in Table 12, the median passing for writing for schools within the first quartile was 35%. This contrasts to a median passing of 68% for students in the fourth quartile. Fourth grade teachers in the first quartile schools correctly judged 67% of students' passing writing. Teachers in fourth quartile schools correctly judged 75% of students' passing writing. Results for the four remaining proficiency tests are listed in Table 12.

Insert Table 12 About Here

Mean score differences in judgments. Means and standard deviations for each judgment condition by grade are presented in Tables 13-to-16. These descriptive statistics are based on the same two schools used in Tables 1-to-10. Mean differences existed for each judgment condition by proficiency area and grade. In all areas, students judged likely to pass had higher mean scores than those judged uncertain to pass. And, except for the fourth grade teachers' judgment of reading, those judged uncertain to pass had higher mean scores than those judged unlikely to pass.

Insert Tables 13-to-16 About Here

ANOVAs. Overall significant differences were found among the judgment categories of likely to pass, uncertain to pass, and unlikely to pass by proficiency area and grade of judgment. Post hoc analyses showed that mean scores associated with likely to pass were significantly greater than unlikely to pass for all proficiency areas and grade levels. In 60% (12) of the judgments, mean scores for likely to pass were significantly greater than uncertain to pass. And in 40% (8) of the judgments, mean scores for uncertain to pass were significantly greater than unlikely to pass. These results are listed in Tables 17-to-20.

Insert Tables 17-to-20 About Here

Discussion

This study provides evidence that teachers can correctly identify many of the students who will pass the Fourth Grade Ohio Proficiency Tests. The median correct judgement of fourth grade teachers in 23 elementary schools ranged from 67% of those students passing science to 81% passing math. However these teachers were less accurate in their judgment of students' failing the Fourth Grade Ohio Proficiency Tests. In this case, the median correct judgement of students' failing ranged from 39% for math to 54% for science. In addition, there was less variability in judging those who passed than in judging those who failed. The semi-interquartile ranges for correct judgment of passing were proportionately much smaller than for correct judgment of failing. This indicates that teachers were more consistent in their judgments of those who passed than those who failed. And for writing, reading, and math, correct judgments related to passing were significantly greater than correct judgments related to failing.

The largest number of judgments came from fourth grade teachers. Fewer judgments were collected from third, fifth, and sixth grade teachers. In general, third and fourth grade teachers showed a high degree of agreement related to students' performance on the Fourth Grade Proficiency Tests. For the school whose data are listed in Tables 1-to-5, fourth grade teachers were more accurate of those students' passing than third grade teachers. In four of the five proficiency areas, third grade teachers were more accurate of those failing.

Tables 6-to-10 list the judgments of fifth and sixth grade teachers in one middle school. In this case, sixth grade teachers were less accurate in all but unlikely to pass science than fifth grade teachers. However, it must be noted that fifth grade teachers only judged 48 of the 156 students judged by sixth grade teachers. Sixth grade teachers only correctly identified 19% of their students who subsequently passed writing, but correctly identified 71% of those who failed writing. Sixth grade teachers were marginally better at identifying those who passed science (38% correct judgments). In the other three proficiency areas, correct judgments for passing ranged from 55% to 68%. Both fifth and sixth grade teachers were more accurate than third and fourth grade teachers in judging those who failed.

In addition to examining teachers' "hits and misses", an analysis of the discrimination of the judgment categories was conducted. Tables 13-to-20 present these findings. The reader is

reminded that these tables are based on the same teachers whose judgments were presented in Tables 1-to-10. The mean performance of students judged likely to pass was greater than those judged uncertain to pass or unlikely to pass regardless of proficiency area or grade of teacher judgment. And in all but fourth grade teacher judgment of reading, the mean performance of students judged uncertain to pass was greater than those judged unlikely to pass. Tests by analysis of variance found that overall mean differences were significantly different. Post hoc analyses showed three significant pairwise differences in 30% of the judgments, two significant pairwise differences in 35% of the judgments, and one significant pairwise difference in 35% of the judgments.

These results show that teachers' judgments are quantitatively different. Each judgment category is often distinct, indicating that teachers are reliably judging student proficiency performance.

At issue now is determining the basis for the teachers' judgments. That is, what might explain the teachers' skill in correctly judging students passing or failing proficiency tests? On the Fourth Grade Proficiency Tests, teachers were more accurate of those passing than failing. Informal follow-up with teachers indicated a reluctance to say a student was going to fail. Although teachers may have believed a student was unlikely to pass, there was the hope that somehow the student would pass. For some teachers, saying a student was unlikely to pass was tantamount to giving up on the student, a sort of "jinxing" the student to fail. However, when a judgment of unlikely to pass was made, fourth grade teachers were rarely wrong.

While sixth grade teachers were generally less accurate than fourth grade teachers in identifying students who passed the various parts of the proficiency tests, they were more accurate in judging those who failed. At this time it is unclear why sixth grade teachers had more difficulty identifying those who passed than identifying those who failed.

At the outset of this study, it was thought that teachers in higher performing schools would be more accurate in their judgments of students than teachers in lower performing schools. To test this, the 23 schools that provided fourth grade proficiency judgments were rank ordered on their students' proficiency results. Schools in the lowest quartile were then compared to schools in the highest quartile. Table 12 shows that schools falling in the first quartile had a median pass rate of 35% for writing, whereas schools in the fourth quartile had a median pass rate of 68%. Teachers in the lowest quartile correctly judged 67% of students passing while

teachers in highest quartile correctly judged 75%. The accuracy of these judgments was not significantly different. This was true for the remaining four proficiency areas, as well.

One possible explanation for teachers' judgment accuracy may rest in teachers' knowledge of the proficiency outcomes. Knowing what students will face on proficiency tests would provide teachers with a basis to judge proficiency performance. To examine this possibility, a follow-up questionnaire was sent to the third to sixth grade teachers (N=90) in the 23 schools asking them to rate the extent to which they know (somewhat well, well, or completely) and teach to (sometimes, frequently, or always) the proficiency outcomes. Fifty-one questionnaires (56.67%) were returned. For the most part, teachers reported that they know the outcomes well-to-very well, and that they frequently-to-always teach to the outcomes.

This is not a surprising finding given the press to align instruction and assessment to the proficiency outcomes. However, although most teachers assert they teach to the outcomes, there are clear differences among schools in student performances, suggesting that teaching effectiveness varies. As well, it is known that teachers' self-reports of teaching practices may not match their actual classroom behaviors (e.g., Stigler & Hiebert, 1997; Witt, 1997).

Judgment as formative assessment. Having established that teachers can reliably predict student performance on Ohio's Fourth and Sixth Grade Proficiency Tests, I am now extending this work to see whether teachers' ability to predict can serve as a formative assessment tool. Formative assessment is diagnostic, and is used to assess strengths and weaknesses in learning, as well as to make changes in the pace or content of instruction (Woolfolk, 1999). Presumably, the ability to predict accurately a future outcome can be used to make repeated judgments in shorter time frames. By doing so, teachers could better assess the efficacy of their interventions and support based on the likely trajectory of students. Repeated judgements of students as uncertain to pass or unlikely to pass should serve to trigger timely teacher and building-level reviews of the quality and kind of help these students need. Ostensibly, as students make progress on the outcomes measured by the proficiency tests, teacher judgments should change accordingly.

I am collecting the repeated judgments of third and fourth grade teachers in two elementary schools. In December, 1999 I started collecting the monthly judgments of 155 students by eight teachers. After teachers rate students I compile the judgments and provide them with the findings and list the students by their judgment status. In February, 2000 I

followed up with the teachers and administrators in both buildings. I am collecting information on how teachers make their judgments, and for those students judged uncertain to pass or unlikely to pass I am recording and collating interventions and other supports provided these at-risk students. Generally, those students judged unlikely to pass were described as lacking sufficient knowledge and skills to pass the proficiency tests. A number of these students were in special education or were being considered for evaluation to determine eligibility for special education. Those judged uncertain to pass were viewed as having the necessary knowledge and skills, but were inconsistent in their performances. These performance deficits were often attributed to motivational, attentional, and the home aspects of the students.

I continued to track the judgments of fourth grade teachers through March, 2000 (the month in which proficiency tests are administered). In June, 2000 I will collect the fourth grade students proficiency results to see the relationship between the teachers' repeated judgments and student performance. I will continue to track the teacher judgments of the third grade cohort through March, 2001. And in June, 2001 I will collect the proficiency results of the current third graders to assess the impact of repeated judgments across grades and over time.

Tables 21 and 22 list the consistency of teacher monthly judgment. In Table 21, for example, 33% of Perry's fourth grade students were consistently judged likely to pass writing. An additional 17% were consistently judged uncertain to pass, and another 22% were judged unlikely to pass. By summing the percentage of students consistently judged likely to pass, uncertain to pass, and unlikely to pass, and then subtracting this total from 100% we find that 28% of students judged monthly had some change in rating. Of those with some change in rating, 9% showed an improvement and 0% showed a decline from the first to the last judgment.

Insert Tables 21-22 About Here

At the heart of school reform is the insistence that all of today's students know more and be able to do more than their counterparts in years past. To accomplish this, standards promoting high performance are now in place. To reach these standards, certain basic practices must be followed. The schools' curricula must be aligned to the standards. Its instructional practices must be aligned to the standards. And assessments must be aligned to the standards. The easiest

of the three to accomplish is curricular alignment. Still most schools in Ohio have yet to align their curriculum to proficiency outcomes. Progress has been made at grade levels where proficiency tests are administered, but less headway is occurring at off-grades. Especially problematic for Ohio schools is the lack of instruction and assessment alignment, particularly at off-grades.

Further study of teacher judgment is necessary, but it offers some promise as part of assessment alignment at the classroom level. Teacher judgment appears to offer useful information on students' likelihood of passing proficiency tests. It is easily gathered and compiled, and may offer sufficient formative feedback to permit teachers to alter instruction and other assistance to deflect a student from failure. Obviously for formative assessment purposes, teacher judgment alone is insufficient. But with other assessments that are validated and linked to proficiency tests, teacher judgment has a meaningful role in supporting student achievement.

References

American Federation of Teachers. (1997, July). Making standards matter 1997. Washington, DC: Author.

Berliner, D. C., & Biddle, B. J. (1995). The manufactured crisis: Myths, fraud, and the attack on America's public schools. New York, NY: Addison-Wesley Publishing.

Bracey, G. W. (1997). The truth about America's schools: The Bracey reports, 1991-1997. Bloomington, IN: Phi Delta Educational Foundation.

Carson, C. C., Huelskamp, R. M., & Woodall, T. D. (1993). Perspectives on education in America: An annotated briefing, April, 1992. The Journal of Educational Research, 86, 259-311.

Coladarci, T. (1986) Accuracy of teacher judgements of student responses to standardized items. Journal of Educational Psychology, 78,141-146.

Demaray, M.K., & Elliott, S.N.(1998). Teachers' judgements of students' academic functioning: A comparison of actual and predicted performances. School Psychology Quarterly, 13, 25-44.

Education Week. (2000, January). Quality counts 2000: Who should teach?. Washington, DC: Author.

Education Week. (1999, January). Quality counts'99: Rewarding results, punishing failure. Washington, DC: Author.

Education Week. (1998, January). Quality counts'98: The urban challenge. Washington, DC: Author.

Education Week. (1997, January). Quality counts: A report card on the condition of public education in the 50 states. Washington, DC: Author.

Education Week on the Web. (1997, October 22). State board's leaders call for assessments bearing consequences. [WWW document]. URL http://www.edweek.com

Forgione, P. D. (1998, April 3). Achievement in the United States: Progress since A Nation at Risk? [WWW document]. URL http://www.nces.ed.gov

Fuller, M. L., & DeMarie-Dreblow, D. (1992, March). Assessing at-risk factors related to performance on Ohio's Ninth Grade Proficiency Tests. Paper presented the 1992 NASP Annual Convention.

Gibson, S., & Dembo, M.H. (1984). Teacher efficacy: A construct validation. Journal of Educational Psychology, 76, 569-582.

Hartman, J. M., & Fuller, M. L. (1997). The development of curriculum-based norms in literature-based classrooms. Journal of School Psychology, 35, 377-389.

Heubert, J. P., & Hauser, R. M. (Eds.). (1999). High stakes testing for tracking, promotion, and graduation. Washington, DC: National Academy Press.

Hoge, R.D., & Coladarci, T. (1989). Teacher-based judgements of academic achievement: A review of literature. Review of Educational Research , 59 297-313.

Hoy, W.K., Woolfolk, A.E. (1993) Teachers' Sense of Efficacy and the Organizational Health of Schools. The Elementary School Journal. 93, 356-372

Levin, H. M. (1998). Educational performance, standards, and the economy. Educational Researcher, 4, 4-10.

Loe, S., & Fuller, M.L. (1997, May). Identifying at-risk factors for poor performance on Ohio's Fourth and Sixth Grade Proficiency Tests. Paper presented at OSPA's annual Spring Conference.

McQuillan, J. (1998). The literacy crisis: False claims, real solutions. Portsmouth, NH: Heinemann.

Ohio Department of Education. (1996, May 14). Proficiency testing in Ohio: A summary. Columbus, OH: Author.

Ohio Department of Education. (1995, August). Fourth-grade proficiency tests: Information guide. Columbus, OH: Author.

Ohio Department of Education. (1995, August). Sixth-grade proficiency tests: Information guide. Columbus, OH: Author.

Robinson, G., & Brandon, D. (1992, September, pp. 30-32). Perceptions about American education: Are they based on facts? Concerns in Education. Arlington, VA: Educational Research Service.

Stanford Achievement Tests: 9th Edition. (1999). Correlation and predictability between the Stanford and the Ohio Proficiencies. San Antonio, TX: Harcourt Brace Educational Measurement.

Stigler, J.A., & Hiebert, J. (1997, September). Understanding and improving classroom mathematics instruction: An overview of the Third International Mathematics and Science Video. Kappan, 14-22.

Tschannen-Moran, M., Hoy, A.W., & Hoy, W.K. (1998) Teacher efficacy: Its meaning and measure. Review of Educational Research, 68, 202-248.

Witt, J.C. (1997). Talk is not cheap. School Psychology Quarterly, 12, 281-292.

Woolfolk, A. E., Rosoff, B, & Hoy, W. K. (1990). Teachers' sense of efficacy and their belief about managing students. Teaching and teacher education. 6, 137-148.

Appendix

Teacher Judgment

District:_____ School:_____

4th Grade Teacher (s):_____ Date:_____

Student Name	Current Grade	Reading	Math	Citizenship	Science	Writing
	4th	LP U ULP	LP U ULP	LP U ULP	LP U ULP	LP U ULP
	4th	LP U ULP	LP U ULP	LP U ULP	LP U ULP	LP U ULP
	4th	LP U ULP	LP U ULP	LP U ULP	LP U ULP	LP U ULP
	4th	LP U ULP	LP U ULP	LP U ULP	LP U ULP	LP U ULP
	4th	LP U ULP	LP U ULP	LP U ULP	LP U ULP	LP U ULP
	4th	LP U ULP	LP U ULP	LP U ULP	LP U ULP	LP U ULP
	4th	LP U ULP	LP U ULP	LP U ULP	LP U ULP	LP U ULP
	4th	LP U ULP	LP U ULP	LP U ULP	LP U ULP	LP U ULP
	4th	LP U ULP	LP U ULP	LP U ULP	LP U ULP	LP U ULP
	4th	LP U ULP	LP U ULP	LP U ULP	LP U ULP	LP U ULP
	4th	LP U ULP	LP U ULP	LP U ULP	LP U ULP	LP U ULP
	4th	LP U ULP	LP U ULP	LP U ULP	LP U ULP	LP U ULP
	4th	LP U ULP	LP U ULP	LP U ULP	LP U ULP	LP U ULP
	4th	LP U ULP	LP U ULP	LP U ULP	LP U ULP	LP U ULP
	4th	LP U ULP	LP U ULP	LP U ULP	LP U ULP	LP U ULP

Table 1

Third and Fourth Grade Teacher Judgment of Student Performance on The Fourth Grade Writing Proficiency Test

		Teacher Judgment		
		% Likely To Pass	% Uncertain	% Unlikey To Pass
	N			
Pass				
Third Grade	54	53.70	33.33	12.96
Fourth Grade	61	80.33	16.39	3.28
Advanced Pass				
Third Grade	5	60.00	40.00	0
Fourth Grade	8	100.00	0	0
Fail				
Third Grade	26	30.77	34.62	34.62
Fourth Grade	27	44.44	37.04	18.52

Note. Pass is writing rubric score of 5 or 6. Advanced pass is 7 or 8.

Table 2

Third and Fourth Grade Teacher Judgment of Student Performance on The Fourth Grade Citizenship Proficiency Test

		Teacher Judgment		
	N	% Likely To Pass	% Uncertain	% Unlikey To Pass
Pass				
Third Grade	69	72.46	23.19	4.35
Fourth Grade	80	86.25	12.50	1.25
Advanced Pass				
Third Grade	9	88.89	11.11	0
Fourth Grade	11	100.00	0	0
Fail				
Third Grade	7	57.14	0	42.86
Fourth Grade	9	22.22	44.44	33.33

Note. Pass is a scaled score of 208 to 249. Advanced pass is 250 or greater.

Table 3

Third and Fourth Grade Teacher Judgment of Student Performance on The Fourth Grade Reading Proficiency Test

		Teacher Judgment		
		% Likely To Pass	% Uncertain	% Unlikey To Pass
	N			
Pass				
Third Grade	68	77.94	16.18	5.88
Fourth Grade	78	84.62	12.82	2.56
Advanced Pass				
Third Grade	2	100.00	0	0
Fourth Grade	2	100.00	0	0
Fail				
Third Grade	15	46.67	33.33	20.00
Fourth Grade	17	35.29	58.82	5.88

Note. Pass is a scaled score of 210 to 249. Advanced pass is 250 or greater.

Table 4

Third and Fourth Grade Teacher Judgment of Student Performance on The Fourth Grade Math Proficiency Test

			Teacher Judgment		
			% Likely To Pass	% Uncertain	% Unlikey To Pass
		N			
Pass					
	Third Grade	57	77.19	19.30	3.51
	Fourth Grade	67	79.10	20.90	0
Advanced Pass					
	Third Grade	10	100.00	0	0
	Fourth Grade	11	100.00	0	0
Fail					
	Third Grade	18	27.79	44.44	27.78
	Fourth Grade	21	33.33	52.38	14.29

Note. Pass is a scaled score of 210 to 249. Advanced pass is 250 or greater.

Table 5

Third and Fourth Grade Teacher Judgment of Student Performance on The Fourth Grade Science Proficiency Test

			Teacher Judgment		
			% Likely To Pass	% Uncertain	% Unlikey To Pass
		N			
Pass					
	Third Grade	43	62.79	34.88	2.33
	Fourth Grade	51	82.35	15.69	1.96
Advanced Pass					
	Third Grade	24	70.83	29.17	0
	Fourth Grade	29	96.55	3.45	0
Fail					
	Third Grade	18	44.44	27.88	27.88
	Fourth Grade	20	30.00	40.00	30.00

Note. Pass is a scaled score of 200 to 249. Advanced pass is 250 or greater.

Table 6

Fifth and Sixth Grade Teacher Judgment of Student Performance on The Sixth Grade Writing Proficiency Test

		Teacher Judgment			
		N	% Likely To Pass	% Uncertain	% Unlikey To Pass
Pass					
	Fifth Grade	16	56.25	25.00	18.75
	Sixth Grade	59	18.64	42.37	38.98
Advanced Pass					
	Fifth Grade	27	74.07	14.82	11.11
	Sixth Grade	80	58.75	28.75	12.50
Fail					
	Fifth Grade	5	20.00	0	80.00
	Sixth Grade	17	0	29.41	70.59

Note. Pass is a writing rubric score of 5 to 6. Advanced pass is 7 to 8.

Table 7

Fifth and Sixth Grade Teacher Judgment of Student Performance on The Sixth Grade Citizenship Proficiency Test

		Teacher Judgment			
		N	% Likely To Pass	% Uncertain	% Unlikey To Pass
Pass					
	Fifth Grade	36	80.56	2.78	16.67
	Sixth Grade	94	55.32	32.98	11.70
Advanced Pass					
	Fifth Grade	5	100.00	0	0
	Sixth Grade	22	90.91	9.09	0
Fail					
	Fifth Grade	7	0	0	100.00
	Sixth Grade	45	0	48.89	51.11

Note. Pass is a scaled score of 200 to 249. Advanced pass is 250 or greater.

Table 8

Fifth and Sixth Grade Teacher Judgment of Student Performance on The Sixth Grade Reading Proficiency Test

		Teacher Judgment		
	N	% Likely To Pass	% Uncertain	% Unlikey To Pass
Pass				
Fifth Grade	27	74.07	18.52	7.41
Sixth Grade	79	60.76	27.85	11.39
Advanced Pass				
Fifth Grade	11	90.91	0	9.09
Sixth Grade	33	87.88	9.09	3.03
Fail				
Fifth Grade	10	20.00	10.00	70.00
Sixth Grade	45	4.44	35.56	60.00

Note. Pass is a scaled score of 211 to 249. Advanced pass is 250 or greater.

Table 9

Fifth and Sixth Grade Teacher Judgment of Student Performance on The Sixth Grade Math Proficiency Test

		Teacher Judgment		
	N	% Likely To Pass	% Uncertain	% Unlikey To Pass
Pass				
Fifth Grade	31	90.32	0	9.68
Sixth Grade	84	67.86	30.95	1.19
Advanced Pass				
Fifth Grade	4	100.00	0	0
Sixth Grade	13	100.00	0	0
Fail				
Fifth Grade	13	23.08	0	76.92
Sixth Grade	62	8.07	48.39	43.55

Note. Pass is a scaled score of 200 to 249. Advanced pass is 250 or greater.

Table 10

Fifth and Sixth Grade Teacher Judgment of Student Performance on The Sixth Grade Science Proficiency Test

		Teacher Judgment		
	N	% Likely To Pass	% Uncertain	% Unlikey To Pass
Pass				
Fifth Grade	34	79.41	8.82	11.77
Sixth Grade	96	37.50	37.50	25.00
Advanced Pass				
Fifth Grade	1	100.00	0	0
Sixth Grade	2	100.00	0	0
Fail				
Fifth Grade	13	23.08	23.08	53.85
Sixth Grade	62	8.07	29.03	62.90

Note. Pass is a scaled score of 200 to 249. Advanced pass is 250 or greater.

Table 11

Median Percent Correct Judgment of Students' Passing or Failing Each Part of OPT

	Pass % Correct	Fail % Correct	Z
Writing	71.9 (10.80)	41.2 (16.10)	-3.528*
Reading	77.0 (11.95)	41.0 (20.15)	-3.523*
Math	81.1 (6.95)	39.2 (21.80)	-3.555*
Citizenship	71.4 (16.90)	40.0 (29.25)	-1.616
Science	66.7 (14.05)	53.8 (15.80)	-1.551

Note. N = 23 schools for all OPT areas. Semi-interquartile ranges are in parentheses. The Wilcoxin Sign Test was used to assess for significance. * $p < .001$.

Table 12

Median Percent Correct Judgment of Students' Passing Each Part of OPT Within Lowest or Highest Quartile Rank

	First Quartile	% Correct	Fourth Quartile	% Correct	Z
Writing	35.25	66.9 (11.6)	67.50	75.4 (6.1)	-.094
Reading	61.00	73.3 (9.3)	78.75	81.6 (16.4)	-.503
Math	43.00	79.4 (3.3)	78.00	84.0 (9.1)	-.656
Citizenship	67.25	57.6 (20.2)	86.00	77.5 (7.2)	-1.403
Science	45.50	64.9 (17.0)	75.75	66.1 (3.3)	-.375

Note. N = 23 schools for all OPT areas. Semi-interquartile ranges are in parentheses. The Mann-Whitney U Test was used to assess for significance. All ps > .05.

Table 13

Means and Standard Deviations of Fourth Grade Proficiency Scores by Third Grade Judgment

	Ohio Fourth Grade Proficiency Tests				
	Writing	Reading	Math	Citizenship	Science
Likely To Pass					
n	40	62	59	62	52
M	5.475	227.177	231.932	232.919	236.962
SD	.905	14.568	20.860	17.699	30.823
Uncertain To Pass					
n	29	16	19	17	27
M	5.034	211.562	215.474	224.882	221.444
SD	1.052	9.633	13.615	11.028	31.100
Unlikely To Pass					
n	15	7	7	6	6
M	4.400	207.143	204.714	208.333	179.333
SD	.910	7.599	7.455	7.312	19.896

Note. Pass scores are: Writing, 5; Reading, 210; Mathematics, 210; Citizenship, 208; and Science, 200.

Table 14

Means and Standard Deviations of Fourth Grade Proficiency Scores by Fourth Grade Judgment

	Ohio Fourth Grade Proficiency Tests				
	Writing	Reading	Math	Citizenship	Science
Likely To Pass					
n	69	74	71	82	76
M	5.464	227.135	232.690	233.732	239.711
SD	.948	13.668	20.035	15.930	27.071
Uncertain To Pass					
n	20	20	25	14	17
M	4.550	208.550	211.600	211.214	199.647
SD	.759	9.633	11.049	8.276	16.507
Unlikely To Pass					
n	6	3	3	4	7
M	4.000	209.333	193.667	204.500	164.571
SD	1.095	8.327	11.372	6.807	25.935

Note. Pass scores are: Writing, 5; Reading, 210; Mathematics, 210; Citizenship, 208; and Science, 200.

Table 15

Means and Standard Deviations of Sixth Grade Proficiency Scores by Fifth Grade Judgment

	Ohio Sixth Grade Proficiency Tests				
	Writing	Reading	Math	Citizenship	Science
Likely To Pass					
N	30	32	35	34	31
M	6.733	240.781	226.800	228.324	219.419
SD	.980	23.965	21.281	17.468	14.787
Uncertain To Pass					
n	8	6	0	1	5
M	6.500	216.667	--	219.00	194.600
SD	.926	9.026	--	---	15.437
Unlikely To Pass					
n	10	10	13	13	11
M	5.300	212.400	186.846	198.538	190.818
SD	1.337	18.887	21.832	22.622	18.983

Note. Pass scores are: Writing, 5; Reading, 211; Mathematics, 200; Citizenship, 200; and Science, 200.

Table 16

Means and Standard Deviations of Sixth Grade Proficiency Scores by Sixth Grade Judgment

	Ohio Sixth Grade Proficiency Tests				
	Writing	Reading	Math	Citizenship	Science
Likely To Pass					
n	58	79	75	72	43
M	7.103	244.165	232.293	235.431	221.558
SD	.693	21.100	20.250	20.867	17.411
Uncertain To Pass					
n	53	41	56	55	54
M	6.264	219.293	200.000	207.564	207.000
SD	1.112	21.756	21.756	19.372	20.600
Unlikely To Pass					
n	45	37	28	34	61
M	5.489	198.595	180.964	191.559	189.230
SD	1.308	25.334	14.574	20.914	22.262

Note. Pass scores are: Writing, 5; Reading, 211; Mathematics, 200; Citizenship, 200; and Science, 200.

Table 17

ANOVA and Post Hoc Results of Student Performances on the Fourth Grade Proficiency Tests as Judged by Third Grade Teachers

	df	F	Scheffé
Writing	2,81	7.07*	Likely To Pass > Unlikely To Pass
Reading	2,82	13.74**	Likely To Pass > Uncertain To Pass Likely To Pass > Unlikely To Pass
Math	2,82	10.43**	Likely To Pass > Uncertain To Pass Likely To Pass > Unlikely To Pass
Citizenship	2,82	7.26**	Likely To Pass > Unlikely To Pass
Science	2,82	10.60**	Likely To Pass > Unlikely To Pass Uncertain To Pass > Unlikely To Pass

$p < .01$ ** $p < .001$

Table 18

ANOVA and Post Hoc Results of Student Performances on the Fourth Grade Proficiency Tests as Judged by Fourth Grade Teachers

	df	F	Scheffé
Writing	2,92	12.87*	Likely To Pass > Uncertain To Pass Likely To Pass > Unlikely To Pass
Reading	2,94	18.85*	Likely To Pass > Uncertain To Pass
Math	2,96	17.64*	Likely To Pass > Uncertain To Pass Likely To Pass > Unlikely To Pass
Citizenship	2,97	19.43*	Likely To Pass > Uncertain To Pass Likely To Pass > Unlikely To Pass
Science	2,97	39.99*	Likely To Pass > Uncertain To Pass Likely To Pass > Unlikely To Pass Uncertain To Pass > Unlikely To Pass

*$p < .001$

Table 19

ANOVA and Post Hoc Results of Student Performances on the Sixth Grade Proficiency Tests as Judged by Fifth Grade Teachers

	df	F	Scheffé
Writing	2,45	6.99*	Likely To Pass > Unlikely To Pass
Reading	2,45	8.11*	Likely To Pass > Unlikely To Pass
Math	1,46	32.96**	Likely To Pass > Unlikely To Pass
Citizenship	2,45	11.58**	Likely To Pass > Unlikely To Pass
Science	2,44	15.80**	Likely To Pass > Uncertain To Pass Likely To Pass > Unlikely To Pass

* $p < .01$ ** $p < .001$

Table 20

ANOVA and Post Hoc Results of Student Performances on the Sixth Grade Proficiency Tests as Judged by Sixth Grade Teachers

	df	F	Scheffé
Writing	2,153	30.56*	Likely To Pass > Uncertain To Pass Likely To Pass > Unlikely To Pass Uncertain To Pass > Unlikely To Pass
Reading	2,154	55.735*	Likely To Pass > Uncertain To Pass Likely To Pass > Unlikely To Pass Uncertain To Pass > Unlikely To Pass
Math	2,156	94.87*	Likely To Pass > Uncertain To Pass Likely To Pass > Unlikely To Pass Uncertain To Pass > Unlikely To Pass
Citizenship	2,158	61.81*	Likely To Pass > Uncertain To Pass Likely To Pass > Unlikely To Pass Uncertain To Pass > Unlikely To Pass
Science	2,155	32.27 *	Likely To Pass > Uncertain To Pass Likely To Pass > Unlikely To Pass Uncertain To Pass > Unlikely To Pass

* $p < .001$

Table 21

Consistency of Monthly Judgments at Perry Elementary by Grade Level

	LTP	UncTP	UnlTP	Change	Improve	Decline
Writing						
3rd Grade	56%	22%	2%	20%	2%	2%
4th Grade	33%	17%	22%	28%	9%	0%
Reading						
3rd Grade	60%	18%	3%	20%	3%	0%
4th Grade	36%	17%	17%	30%	15%	3%
Math						
3rd Grade	58%	15%	3%	24%	6%	3%
4th Grade	28%	19%	8%	45%	26%	3%
Citizenship						
3rd Grade	58%	15%	3%	24%	3%	11%
4th Grade	47%	14%	6%	33%	6%	9%
Science						
3rd Grade	58%	20%	3%	19%	3%	6%
4th Grade	28%	25%	8%	39%	15%	6%

Note. LTP is Likely to Pass. UncTP is Uncertain To Pass and UnlTP is Unlikely To Pass.

Table 22

Consistency of Monthly Judgments at Pike Elementary by Grade Level

	LTP	UncTP	UnlTP	Change	Improve	Decline
Writing						
3rd Grade	17%	28%	10%	45%	29%	0%
4th Grade	28%	25%	8%	39%	22%	5%
Reading						
3rd Grade	38%	14%	7%	41%	10%	0%
4th Grade	38%	18%	0%	44%	27%	0%
Math						
3rd Grade	38%	17%	7%	38%	6%	10%
4th Grade	43%	8%	0%	49%	35%	0%
Citizenship						
3rd Grade	34%	24%	7%	35%	13%	0%
4th Grade	55%	15%	0%	30%	15%	0%
Science						
3rd Grade	14%	45%	7%	34%	9%	0%
4th Grade	13%	20%	15%	52%	24%	3%

Note. LTP is Likely to Pass. UncTP is Uncertain To Pass and UnlTP is Unlikely To Pass.